

Health Disparities in the Medi-Cal Population

Data Sources and Methods



This document provides a summary of the data sources and methods that were used for each of the *Health Disparities in the Medi-Cal Population* fact sheets. The first section provides details about the data sources and methods for fact sheets based on survey data. The second section describes the Medi-Cal Management Information System/Decision Support System (MIS/DSS) data warehouse and how indicators were constructed from fee-for-service claims and managed care encounter data. It also covers the remaining fact sheets that use data sources produced by California government agencies (i.e., Office of Statewide Health Planning and Development (OSHPD), California Department of Public Health (CDPH), and California Department of Education (CDE)). The last section provides details about defining race and ethnicity categories for the fact sheets.

The fact sheets replicate the *Let's Get Healthy California Task Force (LGHCTF) Final Report* in the California Medi-Cal population. The methods and indicators used for the fact sheets have numerous similarities to those used in the *LGHCTF Final Report*. First, it was possible to replicate most of the fact sheets that were based on survey data given the availability of information about Medi-Cal members. Thus, for these indicators, direct comparisons were possible between the overall California rate and the Medi-Cal population. Second, with the exception of the Preventable Hospitalizations fact sheet, rates were not adjusted for age or sex. In addition to the similarities, there were a few notable differences between the fact sheets using non-survey data and some of the indicators in the *LGHCTF Final Report*. Specifically, some of the data were not available to create indicators for the Medi-Cal population. Thus, proxy indicators were created that were conceptually similar, albeit different than the original indicators published in the report. In addition, for some indicators, comparisons had to be made between Medi-Cal and non-Medi-Cal rates rather than the overall California population rate. Finally, in contrast to the survey data with known or limited data quality problems, some of the non-survey indicators had more uncertainty regarding data quality. Details about these issues are found in specific sections below.

Survey Data

Some of the data for these fact sheets were collected from two Computer-Assisted Telephone Interview (CATI) surveys with a similar methodology that allows generalizability to the California state population: 1) California Health Interview Survey (CHIS) and 2) Behavioral Risk Factor Surveillance Survey (BRFSS). Where appropriate, data from both BRFSS and CHIS were weighted to the 2000 Census. Because the age, race, and sex characteristics of respondents differ somewhat from the actual age, race, and sex characteristics of the California population, both surveys used weighting adjustment developed to improve the representativeness of their samples. CHIS was administered to youth aged 12 to 17 and adults over age 18. California BRFSS was administered to adults over the age of 18.

There are limitations when analyzing survey data. There is a certain amount of respondent bias inherent in any study; study participants are usually cooperative and wish to please the interviewer. Data from anonymous and confidential telephone surveys cannot be verified and may be imprecise, especially for more sensitive topics.

The cross-sectional designs of these two surveys also have some inherent limitations. Foremost of these is the inability to determine causation between variables, occurrences, and events. Additionally, while most of the survey scales, indices, and questions were previously validated, the surveys as a whole were not tested for validity or reliability. Following is a description of the surveys and other data sources used for these fact sheets.

California Health Interview Survey

CHIS is the largest state health survey and one of the largest health surveys in the United States. It is a random-digit dial (RDD) telephone survey of households drawn from every county in California. The survey collects information from approximately 50,000 households, and it has been administered every two years since 2001. The CHIS sample is representative of the state's non-institutionalized population living in households. CHIS interviews one sample adult in each household. In households with children, CHIS interviews one adolescent aged 12 to 17 and obtains information for one child under age 12 by interviewing the adult who is most knowledgeable about the child.

The sample size for the 2009 CHIS was 59,938 (47,614 Adults, 8,945 Children, and 3,379 Adolescents). The Medi-Cal sample for CHIS included people who reported having Medi-Cal only or Medi-Cal and Medicare. CHIS covers a wide range of topics, including health status, health conditions, health-related behaviors, health insurance coverage, access to and use of health care services, and the health and development of children and adolescents. To capture the rich diversity of the California population, interviews were conducted in five languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and Korean. These languages were chosen based on analysis of 2000 Census data to identify the languages that would cover the largest number of Californians in the CHIS sample that either did not speak English or did not speak English well enough to otherwise participate.

CHIS is a collaborative project of the University of California, Los Angeles (UCLA) Center for Health Policy Research, CDPH, DHCS, and the Public Health Institute. CHIS is based at the UCLA Center for Health Policy Research in Los Angeles, California. Funding for CHIS comes from state and federal agencies and from several private foundations. Questions and topics in the surveys may vary, dependent on funders' interests. To obtain the data or further information about this survey, interested parties may call (866) 275-2447 or email chis@ucla.edu.

[Adolescent Fruit and Vegetable Consumption, Adolescent Physical Activity, Adolescent Sugar Sweetened Beverages Consumption, Adult Soda and Sweetened Beverages Consumption, Adolescent Obesity, Overall Health Status, and Walking, Biking, and Skating to School Fact Sheets](#)

Data from CHIS were drawn from the public AskCHIS website at: <http://healthpolicy.ucla.edu/chis/Pages/default.aspx>.

[Neighborhood Safety Fact Sheet](#)

CHIS information was not available on the public website (AskCHIS) for the Neighborhood Safety and Overall Health fact sheets; therefore, data were analyzed using public use CHIS data sets.

California Behavioral Risk Factor Surveillance Survey

BRFSS is the world's largest, on-going telephone health survey system, tracking health conditions and risk behaviors among adults over age 18 in the United States annually since 1984. BRFSS provides state-specific information about issues such as diabetes, obesity, cancer screening, nutrition, physical activity, tobacco use, and more.

BRFSS is a state-based system of health surveys that generate information about health risk behaviors, clinical preventive practices, and health care access and use primarily related to chronic diseases and injury. This survey is conducted by the 50 state health departments as well as those in the District of Columbia, Puerto Rico, Guam, and the U.S. Virgin Islands with support from the Centers for Disease Control and Prevention (CDC). BRFSS completes more than 400,000 adult interviews nationally and over 10,000 adult interviews in California each year. California BRFSS is administered in English and Spanish. Response rates measure how successful a survey has been in reaching selected respondents. Two rates are calculated for the BRFSS, an "upper-bound" rate and a CASRO (Council of American Survey Research Organizations) rate. The overall response rate for California BRFSS data used in these fact

sheets ranged from approximately 34% to 58%, depending on the type of response rate calculated (57% to 58% for the upper-bound and 34% to 43% for CASRO). Funding for California BRFSS comes from CDC, California state programs, and several private foundations. Questions and topics in the surveys may vary dependent on funders' interests. To obtain the data or further information about this survey, interested parties may call (916) 779-2677 or email Marta Induni at minduni@s-r-g.org. For more technical information on these surveys, please refer to the BRFSS Documentation and Technical Report.²

Adverse Childhood Experiences, Adult Physical Activity, and Adult Obesity Fact Sheets

During the analyses for the Adverse Childhood Experiences, Adult Physical Activity, and Adult Obesity fact sheets, the California BRFSS was based at the Public Health Institute's Survey Research Group in Sacramento, California. The Medi-Cal program is specific to California; therefore, there was not a Medi-Cal variable in the CDC BRFSS dataset. These fact sheets were focused on comparing the Medi-Cal population to the California population; therefore, the California BRFSS dataset, not the CDC BRFSS dataset, was used in these analyses.

Non Survey Data

Medi-Cal Management Information System/Decision Support System

To manage and store a vast amount of data, DHCS created a data warehouse and reporting system named the MIS/DSS. MIS/DSS contains 10 years of data (about 2.5 billion records) that are extracted from approximately 30 different sources (e.g., eligibility, fee-for-service paid and denied claims, mental health claims, dental claims, and managed care encounter data). In addition, the warehouse includes numerous reference data files to help users map codes to specific labels and descriptions. For example, there are reference tables to The International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9) and other Medi-Cal specific codes. MIS/DSS allows DHCS analytical staff to query specific types of claims or encounters and create analytical reports.

Three main types of Medi-Cal data were used in the fact sheets. First, eligibility data contain records for each month that a potential Medi-Cal member is eligible for services. Dual eligible members (eligible for both Medi-Cal and Medicare) were excluded because Medicare is generally the first payer. These data show when individuals are eligible for Medi-Cal services and thus are useful to construct denominators for rates. Second, there are data related to fee-for-service claims. Fee-for-service claims are submitted by providers to Medi-Cal through a fiscal intermediary for reimbursement for services. Third, managed care encounter data are collected to identify visits and services. Managed care plans are paid on a per member per month basis. Although managed care plans are not paid for individual services, they are required to submit to Medi-Cal "encounter" data for each visit. Fee-for-service claims data are known to be of higher quality in comparison to managed care encounter data given that financial reimbursement is associated with the former. Programs have recently been started, however, to improve the quality of encounter data to ensure that all data are submitted (without duplicates), data elements are correctly coded, and the data represent real health care visits. Although important achievements have been made to improve the quality of encounter data, information derived from these data should be used with caution.

In addition to claims and encounter data, the MIS/DSS system includes numerous tables derived from a product called Symmetry® to "pre-aggregate" claims and encounters into both "episodes of care" and quality indicators. Episode Treatment Groups® (ETG®), which became available in 1993, offer a powerful way of creating episodes of care by placing inpatient, outpatient, and ancillary services into mutually exclusive and exhaustive categories. For example, if a person receives care for a broken arm, the initial visit to treat the break is specified as the "anchor" visit. Then all subsequent follow-up visits and prescriptions are linked to the anchor record to form an episode of care. Although conceptually similar to Diagnosis Related Groups (DRGs), the ETGs identify an entire episode of care regardless of whether it was inpatient or outpatient care. A benefit of using ETGs is that a variety of codes other than diagnosis codes are used to define specific

conditions. For example, many of the ETGs are defined using procedure and revenue codes along with drug codes for specific prescriptions. Given that Medi-Cal only receives two diagnoses as defined by ICD-9 codes, and few validity checks are performed on submitted datasets, it is advantageous that the ETGs define conditions using multiple code types. Concerning quality measurement, the Symmetry® Evidence Based Medicine Connect™ (EBM Connect™) product creates over 580 quality indicators based on the claims and encounter records. Many of the EBM Connect™ quality indicators are based on algorithms that have been validated by groups such as the National Quality Forum.

Adolescent and Adult Depression Fact Sheets

The Adolescent and Adult Depression fact sheets included both fee-for-service claims and managed care encounter data from the MIS/DSS system. The Episode Treatment Groups® system version 8.0 was used to select all claims and encounters associated with episodes of care categorized as “major depression.” The eligibility data from MIS/DSS was used to estimate the number of Medi-Cal members (for the specific age groups) that were at risk of this event.

Hypertension Management, Childhood Immunizations, and Colorectal Cancer Screening Fact Sheets

The three indicators in this group of fact sheets were created using the Symmetry® Evidence Based Medicine Connect™ version 8.0. The EBM Connect™ product implements all of the “business rules” or algorithms that define over 580 quality indicators. Some of the quality indicators are national standards, thus the algorithms copy as closely as possible all of the logic that is required to construct numerators and denominators for rates (e.g., inclusion and exclusion rules). For some of the indicators, EBM Connect™ creates the appropriate cohort of patients based on specific Episode Treatment Groups. For example, the hypertension indicator is generated by first selecting all patients that had an episode of care related to hypertension. There were a few limitations with the indicators in this category. First, since it was not possible to find data sources to replicate the *LGHCTF Final Report* indicators for the Medi-Cal population, results from EBM Connect™ were used to look at health indicators that were conceptually similar to the original indicators. Second, by using the MIS/DSS system for these indicators data were unavailable to report comparison rates for the non-Medi-Cal population. Third, all of the indicators described here used both fee-for-service claims and managed care encounter data. Given the known limitations of managed care encounter data, the results of these four fact sheets should be viewed as preliminary until additional validation has been performed or better data and associated indicators become available. Finally, the hypertension indicator has not been validated by expert panels.

Hypertension Management. There were four separate indicators in this fact sheet to evaluate if disparities existed with regards to how well Medi-Cal members from different racial/ethnic groups managed their hypertension. These medication adherence indicators relied on pharmacy claims data and were created to determine the degree to which a patient was adhering to a prescribed medication based on prescription filling patterns. Technical staff working on validating the encounter data has found that the pharmacy claims data were generally well reported and more reliable than the medical encounters.

Childhood Immunizations. The reported childhood immunization indicators from EBM Connect™ were similar but not identical to the national standard indicators on which they were based. These indicators could not be replicated exactly because the original indicators involved a hybrid methodology of both administrative and clinical data.

Colorectal Cancer Screening. The EBM Connect™ system used a nationally validated Healthcare Effectiveness Data and Information Set (HEDIS) measure to create this colorectal cancer screening indicator. HEDIS is a set of nationally recognized performance indicators used to evaluate health plans. HEDIS national standards are based on selected HEDIS indicators that assess compliance with chronic and acute conditions, as well as preventive care.

Office of Statewide Health Planning and Development

Two datasets available from OSHPD were used for specific fact sheets listed below. First, California licensed hospitals are required to submit data to OSHPD every six months about patients that are discharged from their facilities. The Patient Discharge Dataset (PDD) includes elements related to demographics (e.g., age, gender, race/ethnicity), diagnoses and procedures, expected source of payment, total charges, length-of-stay, and additional fields related to source of admission and place of discharge. The PDD is subjected to validation rules upon submission to OSHPD and has a reputation of being a useful and high-quality dataset. Second, beginning in 2005, OSHPD began collecting encounter records associated with patients who are treated in licensed California emergency departments. Similar to the Patient Discharge Data, the Emergency Department (ED) dataset contains data elements related to demographics, treatment information, patient disposition, and expected source of payment. More information can be found on the OSHPD website: <http://www.oshpd.ca.gov/HID/Products>.

Analytical staff in the OSHPD Healthcare Information Division and the CDPH California Breathing unit produced information for the fact sheets below.

Hospital-Acquired Conditions Fact Sheet

The Agency for Healthcare Research and Quality (AHRQ) releases software each year to produce a variety of quality indicators using administrative hospital data. The Patient Safety Indicators (PSIs) are a set of indicators that can be used to identify potential adverse events that occur after procedures and childbirth. More information about the indicators can be found on the AHRQ website: http://www.qualityindicators.ahrq.gov/modules/psi_overview.aspx.

OSHPD staff applied the SAS software version 4.4 of the PSIs to the 2011 Patient Discharge Data (PDD). Their analysis involved a few steps. First, the PDD was modified to conform to the data specifications required to produce the PSIs. Second, important parameters were set in the AHRQ SAS software to conform to the input data, and the PSI algorithms were applied to transform patient-level data into hospital observed and risk-adjusted rates. Third, a separate SAS program was run to produce a composite PSI indicator. The PSI composite combined 11 of the individual PSIs (using various weights) into one global measure of patient safety. Finally, using an “expected payer” field available in the PDD, the OSHPD researchers stratified the composite measure by Medi-Cal and non-Medi-Cal. It is important to note that the expected payer field is known to have data reliability issues, and thus the results should be compared to other data sources. For example, a preliminary file was used to link the OSHPD patient discharge dataset to the Medi-Cal eligibility file. Using only records that matched on Social Security Number, problems were identified with the expected payer field. Staff is currently working on a more sophisticated probabilistic linkage, and when complete, this file can be used to create an updated fact sheet. For the current fact sheet, however, it is likely that the “expected payer” field is robust enough to present preliminary results.

Preventable Hospitalizations Fact Sheet

In addition to the PSIs, AHRQ also produced a set of indicators to measure “ambulatory care sensitive conditions.” These conditions are chronic and acute conditions that if properly managed by patients and primary care physicians, are less likely to lead to acute complications that require hospitalizations. The AHRQ Prevention Quality Indicators (PQIs) are available to measure a number of chronic and acute conditions, and a global composite measure is provided to weight and aggregate a number of the individual indicators. OSHPD research staff followed similar steps as described above for the PSIs to produce PQI numerators.

The PQIs were created as “area” indicators—numerators for specific geographic areas such as counties were divided by population data to get rates per 100,000 people. For example, OSHPD produced county-level PQI reports that showed the likelihood that people from particular counties were hospitalized for chronic and acute conditions. To produce the PQI composite measure for the Medi-Cal population, staff could not rely on the published census data that were integrated into the AHRQ software. Thus, DHCS staff received

the PQI numerator data (stratified by race/ethnicity) from OSHPD staff, and then relied on eligibility data from the Medi-Cal MIS/DSS system (described above) to produce the denominators. For each racial/ethnic group, DHCS staff excluded “dual eligible” members that were eligible for both Medi-Cal and Medicare. Since Medicare is the primary payer for this population, Medi-Cal has incomplete claims/encounter data for this group.

Hospital Readmissions Fact Sheet

OSHPD research staff received from the Centers for Medicare and Medicaid Services (CMS) a preliminary version of the SAS computer code used to produce hospital readmission rates (the same staff also produced the rate for the *LGHCTF Final Report*). Staff applied the algorithms to the PDD to identify numerators and denominators. Similar to other analyses using the PDD described above, staff used the “expected payer” field to identify Medi-Cal and non-Medical patients (see note above about data quality issues associated with this field).

Childhood Asthma Emergency Department Visits Fact Sheet

OSHPD ED and PDD data were used to create numerators for the childhood asthma rate. Researchers from CDPH counted ED visits among children and adolescents aged 0 to 17 with a principal diagnosis defined by the ICD-9 code 493. When an ED visit resulted in a hospital admission, the visit was only recorded in the Patient Discharge Dataset. Thus, it was necessary to add visits from the hospital discharge database to the counts of ED visits (inpatient discharges that did not include an ED visit were excluded). With this methodology, the numerator counts were based on the number of ED visits, and not the number of unique individuals. Finally, the expected payer field within the OSHPD ED dataset was used to identify Medi-Cal members.

After receiving the stratified counts by race/ethnicity, the Medi-Cal MIS/DSS system was used to estimate the specific denominators to construct the rates. Medi-Cal members who were eligible for both Medi-Cal and Medicare (“dual eligibles”) were excluded given that Medicare is the primary payer for this population and thus there are incomplete claims and encounter data for this group.

Palliative Care Fact Sheet

Using OSHPD Patient Discharge Data, researchers at the University of California, San Francisco identified 351 California acute care hospitals that they expected to have a palliative care program. They submitted a web-based survey with branching logic to each hospital to learn if they had a palliative care program. If any palliative care services were offered, a hospital was designated as a palliative care hospital. OSHPD publishes inpatient profiles on their website. These reports used the “expected payer” source from the Patient Discharge Data to count the number of hospital discharges with an expected payer of Medi-Cal. Looking at all of the general acute care hospitals in California, hospitals were categorized by the number of Medi-Cal members treated in 2011. Dr. Steven Z. Pantilat’s team at the University of California, San Francisco linked their palliative care survey data to the list of hospitals categorized by Medi-Cal patients to identify the degree to which hospitals with a larger proportion of Medi-Cal members had a palliative care program.

California Department of Education

CDE publishes two data sources important for two fact sheets. First, the California Education Code since 1996 has mandated educational agencies to administer a physical fitness test (PFT) to all fifth, seventh, and ninth graders. The State Board of Education selected a test called the Fitnessgram® to measure minimum fitness levels that are likely associated with characteristics that can prevent inactivity-related diseases. Second, each spring students in grades two through eleven must take a Standardized Testing and Reporting (STAR) test. The purpose of the STAR tests is to assess how well students are doing in various subjects such as math, reading, writing, science, and history. Both the STAR test results and the PFT data can be queried from the DataQuest system on the CDE website: <http://data1.cde.ca.gov/dataquest/>.

Reading Proficiency and Childhood Physical Fitness Fact Sheets

At the time of publication, there were no Medi-Cal specific data available to measure childhood physical fitness or reading proficiency. However, PFT and STAR data can be stratified by socio-economic status. Students were classified as “economically disadvantaged” if they participated in the free or reduced price meal program, or if their parent education level was coded as “not high school graduate.” Within the DataQuest query system, the economically disadvantaged variable was used as a proxy measure for the Medi-Cal population given that the vast majority of Medi-Cal members are near or below federal poverty boundaries. In addition, the data were stratified by race/ethnicity.

Birth Cohort File

Infant Mortality Fact Sheet

Each year, analytical staff from CDPH link vital statistics from birth and death files to create the Birth Cohort File. The file allows researchers to evaluate both infant deaths and birth outcomes, and represents all of the live births that occurred in California for the calendar year. In addition, death information is available for all infants born in the calendar year but who died within 12 months of birth. More information can be found on the CDPH website: <http://www.cdph.ca.gov/data/dataresources/requests/Pages/BirthandFetalDeathFiles.aspx>.

Using the Birth Cohort File, staff from CDPH created an infant mortality rate. The numerator was created by selecting infant deaths where the source of payment for prenatal care or delivery is indicated as Medi-Cal (or non-Medi-Cal). The denominator included all live births to California state residents where the source of payment for prenatal care or delivery is indicated as Medi-Cal (or non-Medi-Cal). Using the Birth Cohort File, the overall California infant mortality rate was 4.9 per 1,000 births. This is higher than the overall rate of 4.7 per 1,000 births that is reported on the CDPH website. The CDPH report has a lower infant mortality rate because this report used an unlinked death file that had fewer reported deaths than the Birth Cohort File. The Birth Cohort File is produced after the death file and thus is able to incorporate out-of-state deaths and any additional deaths not reported in the original death file.

To stratify by the Medi-Cal population, two variables from the Birth Cohort File were used. First, two data elements were selected from the data field labeled “principle source of payment for prenatal care”: “02” Medi-Cal, without Comprehensive Perinatal Services Program (CPSP) Support Services or “13” Medi-Cal, with CPSP Support Services. Second, one data element was selected from the field labeled “expected principle source of payment for delivery”: “02” Medi-Cal. With detailed demographic data on the linked file, CDPH staff was able to stratify the results by race/ethnicity for the Medi-Cal and non-Medi-Cal population.

Defining Race/Ethnicity Categories

Although attempts were made to standardize terms and nomenclature, some variability was unavoidable due to the nature of the different data sources. For example, the OSHPD Patient Discharge Data and Emergency Department data have an ethnicity field for Hispanic and Non-Hispanic and a race field that includes White, African American, Native American, Asian/Pacific Islander, Other, and Unknown. The Hispanic field was re-labeled “Latino” and in some instances was used in conjunction with race to form an “Other/Non-Latino” category. The eligibility data from the (MIS/DSS) data warehouse have detailed racial/ethnic breakdowns (e.g., Hmong, Vietnamese). To maintain consistency with other fact sheets, Asian categories were aggregated and included with Pacific Islanders. Finally, the MIS/DSS has missing data for about 10% of the eligible members. These members were grouped with the “Other” category to form an “Other/Unknown” category. The vital statistics from CDPH included similar ethnic/racial categories. The fact sheet using these data also created an “Other/Unknown” category. The survey data sources have similar, yet slightly different categories. CHIS reported race/ethnicity categories for Latino, African American, White, Asian/Other, and Multiracial, while BRFSS only allowed race/ethnicity categories for four groups (Latino,

African American, White, and Asian/Other). In some instances, the survey sample size for the Medi-Cal population was too small to analyze race/ethnicity differences; therefore, in these cases, gender differences were examined.

1. Let's Get Healthy California Task Force Final Report. <http://www.chhs.ca.gov/Documents/Let%27s%20Get%20Healthy%20California%20Task%20Force%20Final%20Report.pdf>. Published December 19, 2013. Accessed February 25, 2013.
2. California Behavioral Risk Factor Survey SAS Dataset Documentation and Technical Report, Survey Research Group, California Department of Health Services, 2006.